

### **3. METHODS USED TO DEVELOP THE 2007 GUIDELINE**

#### **3.1 OVERVIEW**

The development of this guideline drew upon methods outlined by NICE (*Guideline Development Methods: Information for National Collaborating Centres and Guideline Developers*<sup>2</sup> [NICE, 2005b]). A team of health professionals, lay representatives and technical experts known as the GDG, with support from the NCCMH staff, undertook the development of a patient-centred, evidence-based guideline. There are six basic steps in the process of developing a guideline:

- define the scope, which sets the parameters of the guideline and provides a focus and steer for the development work
- define clinical questions considered important for practitioners and service users
- develop criteria for evidence searching and search for evidence
- design validated protocols for systematic review and apply to evidence recovered by search
- synthesise and (meta-) analyse data retrieved, guided by the clinical questions, and produce evidence profiles
- answer clinical questions with evidence-based recommendations for clinical practice.

The clinical practice recommendations made by the GDG are therefore derived from the most up-to-date and robust evidence for the clinical and cost effectiveness of treatments and services used in the management of mental health disorders in women during pregnancy and up to 1 year after delivery. In addition, to ensure a service user and carer focus, the concerns of service users and carers regarding clinical practice have been highlighted and addressed by recommendations agreed by the whole GDG.

#### **3.2 THE SCOPE**

Guideline topics are selected by the Department of Health (DH) and the Welsh Assembly Government, which identify the main areas to be covered by the guideline in a specific remit (see *The Guideline Development Process – An Overview for Stakeholders, the Public and the NHS*<sup>3</sup> [NICE, 2004e]). The remit for this guideline was translated into a scope document by staff at the NCCMH.

---

<sup>2</sup>Available from: [www.nice.org.uk](http://www.nice.org.uk)

<sup>3</sup>Available from: [www.nice.org.uk](http://www.nice.org.uk)

## *Methods used to develop this guideline*

The purpose of the scope was to:

- provide an overview of what the guideline would include and exclude
- identify the key aspects of care that must be included
- set the boundaries of the development work and provide a clear framework to enable work to stay within the priorities agreed by NICE and the NCCMH and the remit from the DH/Welsh Assembly Government
- inform the development of the clinical questions and search strategy
- inform professionals and the public about the expected content of the guideline
- keep the guideline to a reasonable size to ensure that its development could be carried out within an 18-month period.

The draft scope was subject to consultation with stakeholders over a 4-week period. During the consultation period, the scope was posted on the NICE website ([www.nice.org.uk](http://www.nice.org.uk)). Comments were invited from stakeholder organisations and the Guideline Review Panel (GRP). Further information about the GRP can also be found on the NICE website. The NCCMH and NICE reviewed the scope in light of comments received, and the revised scope was signed off by the GRP.

### **3.3 THE GUIDELINE DEVELOPMENT GROUP**

The GDG was made up of professionals in psychiatry, clinical psychology, midwifery, health visiting, social work and general practice, together with two former service users. The guideline development process was supported by staff from the NCCMH, who undertook the clinical and health economics literature searches, reviewed and presented the evidence to the GDG, managed the process and contributed to drafting the guideline.

#### **3.3.1 Guideline Development Group meetings**

Fifteen GDG meetings were held between 18 November 2004 and 29 September 2006. During each day-long GDG meeting, in a plenary session, clinical questions and clinical and economic evidence were reviewed and assessed, and recommendations formulated. At each meeting, all GDG members declared any potential conflict of interest, and service-user concerns were routinely discussed as part of a standing agenda.

#### **3.3.2 Topic groups**

The GDG divided its workload along clinically relevant lines to simplify the guideline development process, and GDG members formed smaller topic groups to undertake guideline work in that area of clinical practice. Topic Group 1 covered questions relating to pharmacological aspects of management of antenatal and postnatal mental health problems; Topic Group 2 covered the prediction and detection of mental disorder; Topic

Group 3 covered psychology and psychosocial interventions; and Topic Group 4 covered service delivery. These groups were designed to efficiently manage the large volume of evidence appraisal prior to presenting it to the GDG as a whole. Each topic group was chaired by a GDG member with expert knowledge of the topic area (one of the healthcare professionals). Topic groups refined the clinical questions, refined the clinical definitions of treatments, reviewed and prepared the evidence with the systematic reviewer before presenting it to the GDG as a whole and helped the GDG to identify further expertise in the topic. Topic-group leaders reported the status of the group's work as part of the standing agenda. They also introduced and led the GDG discussion of the evidence review for that topic and assisted the GDG Chair in drafting that section of the guideline relevant to the work of each topic group.

### **3.3.3 Service users and carers**

Individuals with direct experience of services gave an integral service-user focus to the GDG and the guideline. The GDG included two service users. They contributed as full GDG members to writing the clinical questions, helping to ensure that the evidence addressed their views and preferences, highlighting sensitive issues and terminology relevant to the guideline and bringing service-user research to the attention of the GDG. In drafting the guideline, they contributed to identifying recommendations from the service-user perspective. In addition, testimonies were collected from other service users and healthcare professionals (see Section 3.8).

### **3.3.4 Special advisers**

Special advisers, who had specific expertise in one or more aspects of treatment and management relevant to the guideline, assisted the GDG, commenting on specific aspects of developing the guideline and making presentations to the GDG. Appendix 2 lists those who agreed to act as special advisors.

### **3.3.5 Consensus conference and focus group**

A consensus conference was held during the guideline development period in collaboration with the GDG developing the NICE guideline for the treatment and management of bipolar disorder. This was to discuss the use of psychotropic medication before, during and after pregnancy with invited experts from outside of the GDG, who gave presentations and commented on a draft position statement which formed the basis of Chapter 7. Invited experts are listed in Appendix 2.

Towards the end of the guideline development process, a focus group was held with healthcare professionals from primary care (GPs, health visitors and midwives) to aid understanding of how the guideline will impact on primary care in order to facilitate writing the quick reference guide (see Section 2.2.4).

### 3.4 CLINICAL QUESTIONS

Clinical questions were used to guide the identification and interrogation of the evidence base relevant to the topic of the guideline. Before the first GDG meeting, draft questions were prepared by NCCMH staff based on the scope and an overview of existing guidelines and modified during a meeting with the guideline Chair. They were then discussed by the GDG and amended as necessary. Where appropriate, the questions were refined once the evidence had been searched and, where necessary, sub-questions were generated. Questions submitted by stakeholders were also discussed by the GDG and the rationale for not including questions was recorded in the minutes. The final list of clinical questions is in Appendix 5.

For questions about interventions, the patient, intervention, comparison and outcome (PICO) framework was used. This structured approach divides each question into four components: the patients (the population under study), the interventions (what is being done), the comparisons (other main treatment options) and the outcomes (the measures of how effective the interventions have been) (see Text Box 1).

For questions relating to diagnosis, the PICO framework was not used, as such questions do not involve an intervention designed to treat a particular condition. Rather, the questions were designed to pick up key issues specifically relevant to diagnostic tests, for example their accuracy, reliability, safety and acceptability to the patient.

#### **Text Box 1: Features of a well-formulated question on intervention effectiveness – the PICO guide**

<b>Patients/population</b>	Which patients or population of patients are we interested in? How can they be best described? Are there subgroups that need to be considered?
<b>Intervention</b>	Which intervention, treatment or approach should be used?
<b>Comparison</b>	What is/are the main alternative/s to compare with the intervention?
<b>Outcome</b>	What is really important for the patient? Which outcomes should be considered: intermediate or short-term measures, mortality, morbidity and treatment complications, rates of relapse, late morbidity and readmission, return to work, physical and social functioning and other measures such as quality of life, general health status and costs?

In some situations, the prognosis of a particular condition is of fundamental importance over and above its general significance in relation to specific interventions. Areas where this is particularly likely to occur relate to assessment of risk, for example in terms of behaviour modification or screening and early intervention. In addition, questions related to issues of service delivery are occasionally specified in the remit from the DH/Welsh Assembly Government. In these cases, appropriate clinical questions were developed to be clear and concise.

To help facilitate the literature review, a note was made of the best study-design type to answer each question. There are four main types of clinical questions of relevance to NICE guidelines. These are listed in Text Box 2. For each type of question, the best primary study design varies, where 'best' is interpreted as 'least likely to give misleading answers to the question'.

However, in all cases, a well-conducted systematic review of the appropriate type of study is likely to always yield a better answer than a single study.

Deciding on the best design type to answer a specific clinical or public health question does not mean that studies of different design types addressing the same question were discarded.

**Text Box 2: Best study design to answer each type of question**

<b>Type of question</b>	<b>Best primary study design</b>
Effectiveness or other impact of an intervention	Randomised controlled trial (RCT); other studies that may be considered in the absence of an RCT are the following: internally/externally controlled before-and-after trial, interrupted time series
Accuracy of information (for example, risk factor, test, prediction rule)	Comparing the information against a valid gold standard in a randomised trial or inception cohort study
Rates (of disease, patient experience, rare side effects)	Cohort, registry, cross-sectional study
Costs	Naturalistic prospective cost study

### **3.5 SYSTEMATIC CLINICAL LITERATURE REVIEW**

The aim of the clinical literature review was to systematically identify and synthesise relevant evidence from the literature in order to answer the specific clinical questions developed by the GDG. Thus, clinical practice recommendations are evidence based, where possible, and if evidence was not available, informal consensus methods were used (see Section 3.5.6) and the need for future research was specified.

### **3.5.1 Methodology**

A stepwise, hierarchical approach was taken to locating and presenting evidence to the GDG. The NCCMH developed this process based on methods set out in *Guideline Development Methods: Information for National Collaborating Centres and Guideline Developers*<sup>4</sup> (NICE, 2005b) and after considering recommendations from a range of other sources. These included:

- Centre for Clinical Policy and Practice of the New South Wales Health Department (Australia)
- Clinical Evidence
- The Cochrane Collaboration
- New Zealand Guidelines Group
- NHS Centre for Reviews and Dissemination
- Oxford Centre for Evidence-Based Medicine
- Scottish Intercollegiate Guidelines Network (SIGN)
- United States Agency for Healthcare Research and Quality
- Oxford Systematic Review Development Programme
- GRADE Working Group.

### **3.5.2 The review process**

After the scope was finalised, a more extensive search for systematic reviews and published guidelines was undertaken.

The GDG decided which questions were likely to have a good evidence base and which questions were likely to have little or no directly relevant evidence. In the absence of good evidence, recommendations were developed by informal consensus. For questions that were unlikely to have a good evidence base, a brief descriptive review was initially undertaken by a member of the GDG (see Section 3.5.6). For questions with a good evidence base, the review process depended on the type of clinical question.

Searches for evidence were updated between 6 and 8 weeks before the first consultation. After this point, studies were included only if they were judged by the GDG to be exceptional (for example, the evidence was likely to change a recommendation).

#### *The search process for questions concerning interventions*

For questions related to interventions, the initial evidence base was formed from well-conducted RCTs that addressed at least one of the clinical questions. Although there are a number of difficulties with the use of RCTs in the evaluation of interventions in mental health, the RCT remains the most important method for establishing treatment efficacy (this is discussed in more detail in appropriate clinical evidence chapters). For other clinical questions, searches were for the appropriate study design (see above).

---

<sup>4</sup>Available from: [www.nice.org.uk](http://www.nice.org.uk)

All searches were based on the standard mental-health-related bibliographic databases (EMBASE, MEDLINE, CINAHL, PsycINFO) for all trials potentially relevant to the guideline. Since the number of citations generated from a search for all RCTs was large (around 14,000), this search was run three times: once for citations up to 1994, once for citations from 1995 to 1999 and a third for citations from 2000 to 2004 (when the development process started). Update searches were undertaken a further two times during the development process. Additional searches were run for clinical questions not best answered by RCTs. These are noted in the review write-ups in the following chapters.

After the initial search results were scanned liberally to exclude irrelevant papers, the review team used a purpose-built 'study information' database to manage both the included and the excluded studies (eligibility criteria were developed after consultation with the GDG). Future guidelines will be able to update and extend the usable evidence base starting from the evidence collected, synthesised and analysed for this guideline.

In addition, searches were made of the reference lists of existing systematic reviews and included studies, as well as the list of evidence submitted by stakeholders. Known experts in the field, based both on the references identified in early steps and on advice from GDG members, were sent letters requesting relevant studies that were in the process of being published<sup>5</sup>. In addition, the tables of contents of appropriate journals were periodically checked during the development process for relevant studies.

### *Search filters*

Search filters developed by the review team consisted of a combination of subject heading and free-text phrases. Specific filters were developed for the guideline topic and, where necessary, for each clinical question. In addition, the review team used filters developed for systematic reviews, RCTs and other appropriate research designs (see Appendix 6).

### *Study selection*

All primary-level studies included after the first scan of citations were acquired in full and re-evaluated for eligibility at the time they were being entered into the study information database. Appendix 7 lists the standard inclusion criteria. More specific eligibility criteria were developed for each clinical question and are described in the relevant clinical evidence chapters. Eligible primary-level studies were critically appraised for methodological quality (see Appendix 8). The eligibility of each study was confirmed by at least one member of the appropriate topic group.

For some clinical questions, it was necessary to prioritise the evidence with respect to the UK context (that is, external validity). To make this process explicit, the

---

<sup>5</sup>Unpublished full trial reports were also accepted where sufficient information was available to judge eligibility and quality (see section on unpublished evidence).

### *Methods used to develop this guideline*

topic groups took into account the following factors when assessing the evidence:

- participant factors (for example, gender, age and ethnicity)
- provider factors (for example, model fidelity, the conditions under which the intervention was performed and the availability of experienced staff to undertake the procedure)
- cultural factors (for example, differences in standard care and differences in the welfare system).

It was the responsibility of each topic group to decide which prioritisation factors were relevant to each clinical question in light of the UK context, and then decide how they should modify their recommendations.

### *Unpublished evidence*

The GDG used a number of criteria when deciding whether or not to accept unpublished data. First, the evidence must have been accompanied by a trial report containing sufficient detail to properly assess the quality of the data. Second, the evidence must have been submitted with the understanding that data from the study and a summary of the study's characteristics would be published in the full guideline. Therefore, the GDG did not accept evidence submitted as commercial in confidence. However, the GDG recognised that unpublished evidence submitted by investigators might later be retracted by those investigators if the inclusion of such data would jeopardise publication of their research.

### **3.5.3 Synthesising the evidence**

Outcome data were extracted from all eligible studies that met the quality criteria, using standardised forms (see Appendix 9 and Appendix 10). Where possible, meta-analysis was used to synthesise the evidence using Review Manager 4.2.8 (Cochrane Collaboration, 2005). If necessary, reanalyses of the data or sub-analyses were used to answer clinical questions not addressed in the original studies or reviews.

For a given outcome (continuous and dichotomous), where more than 50% of the number randomised to any group were not accounted for<sup>6</sup> by trial authors, the data were excluded from the review because of the risk of bias. However, where possible, dichotomous efficacy outcomes were calculated on an intention-to-treat basis (that is, a 'once-randomised-always-analyse' basis). This assumes that those participants who ceased to engage in the study – from whatever group – had an unfavourable outcome and means that the 50% rule was not applied to dichotomous outcomes where there was good evidence that those participants who ceased to engage in the study were likely to have an unfavourable outcome (in this case, early withdrawals were included in both the numerator and denominator). Adverse effects were entered into Review Manager as reported by the study authors because it was usually not possible to

---

<sup>6</sup>'Accounted for' in this context means that an appropriate method for dealing with missing data (for example, last observation carried forward [LOCF] or a regression technique) had been used.



determine whether early withdrawals had an unfavourable outcome. For the outcome 'leaving the study early for any reason', the denominator was the number randomised.

The number needed to treat to benefit (NNTB) or the number needed to treat to harm (NNTH) was reported for each outcome where the baseline risk (that is, control group event rate) was similar across studies. In addition, NNTs calculated at follow-up were only reported where the length of follow-up was similar across studies. When the length of follow-up or baseline risk varies (especially with low risk), the NNT is a poor summary of the treatment effect (Deeks, 2002).

Included/excluded studies tables, generated automatically from the study information database, were used to summarise general information about each study (see Appendix 18). Where meta-analysis was not appropriate and/or possible, the reported results from each primary-level study were also presented in the included studies table (and included, where appropriate, in a narrative review).

Consultation was used to overcome difficulties with coding. Data from studies included in existing systematic reviews were extracted independently by one reviewer and cross-checked with the existing data set. Where possible, two independent reviewers extracted data from new studies. Where double data extraction was not possible, data extracted by one reviewer was checked by the second reviewer. Disagreements were resolved with discussion. Where consensus could not be reached, a third reviewer resolved the disagreement. Masked assessment (that is, blind to the journal from which the article comes, the authors, the institution and the magnitude of the effect) was not used since it is unclear that doing so reduces bias (Jadad *et al.*, 1996; Berlin, 1997).

### **3.5.4 Presenting the data to the GDG**

Summary characteristics tables and, where appropriate, forest plots generated with Review Manager, were presented to the GDG, in order to prepare an evidence profile for each review and to develop recommendations.

#### *Evidence profile tables*

An evidence profile table was used to summarise both the quality of the evidence and the results of the evidence synthesis (see Table 1 for an example evidence profile table). Each table included details about the quality assessment of each outcome: number of studies, the study design, limitations (based on the quality of individual studies; see Appendix 8 for the quality checklist and Appendix 18 for details about each study), information about the consistency of the evidence (see below for how consistency was measured), directness of the evidence (that is, how closely the outcome measures, interventions and participants match those of interest) and any other considerations (for example, effect sizes with wide confidence intervals (CIs) would be described as imprecise data). Each evidence profile also included a summary of the findings: number of patients included in each group, an estimate of the magnitude of the effect, quality of the evidence and the importance of the evidence. The quality of the evidence was based on the quality assessment

Table 1: Example evidence profile table

Quality assessment					Summary of findings					
No. of studies	Quality of included studies	Consistency	Directness	Other modifying factors	No. of patients		Effect		Overall quality	Importance
					CBT	Waitlist	Effect size	Likelihood of clinically important effect		
<b>Carbamazepine versus placebo</b>										
<b>Efficacy</b>										
Mania: change score at endpoint (manic only)										
Weisler 2004	Sign: 1 + (-1)	None	Acute mania	Strong effect (+1)	134	146	WMD = -6.87 (-9.24, -4.49)	Very likely (favouring carbamazepine)	Moderate	Critical
Weisler 2005										
<b>Acceptability/tolerability</b>										
Leaving treatment early for any reason										
Weisler 2004	Sign: 1 + (-1)	Significant heterogeneity (-1): random effects model used	Acute mania	None	134	146	RR (random effects) = 0.79 (0.59, 1.06)	Likely (favouring carbamazepine)	Low	Critical
Weisler 2005										

components (study design, limitations to study quality, consistency, directness and any other considerations) and graded using the following definitions:

- **High** = Further research is very unlikely to change our confidence in the estimate of the effect.
- **Moderate** = Further research is likely to have an important impact on our confidence in the estimate of the effect and may change the estimate.
- **Low** = Further research is very likely to have an important impact on our confidence in the estimate of the effect and is likely to change the estimate.
- **Very low** = Any estimate of effect is very uncertain.

For further information about the process and the rationale of producing an evidence profile table, see GRADE Working Group (2004).

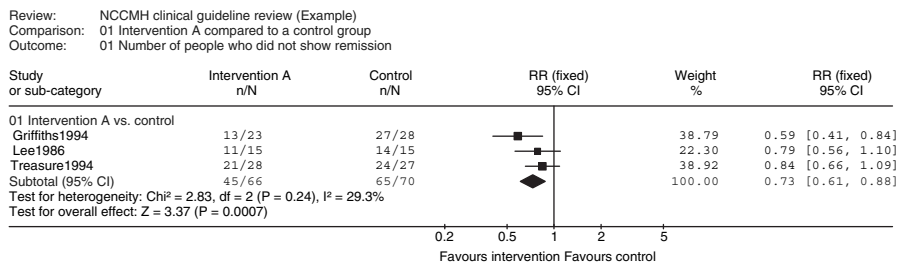
*Forest plots*

Each forest plot displayed the effect size and CI for each study as well as the overall summary statistic. The graphs were organised so that the display of data in the area to the left of the ‘line of no effect’ indicated a ‘favourable’ outcome for the treatment in question. Dichotomous outcomes were presented as relative risks with the associated 95% CI (for an example, see Figure 1). A relative risk (or risk ratio) is the ratio of the treatment event rate to the control event rate. A relative risk of 1 indicates no difference between treatment and control. In Figure 1, the overall relative risk of 0.73 indicates that the event rate (that is, non-remission rate) associated with intervention A is about three quarters of that with the control intervention, or in other words, the relative risk reduction is 27% (that is, 270 in 1,000).

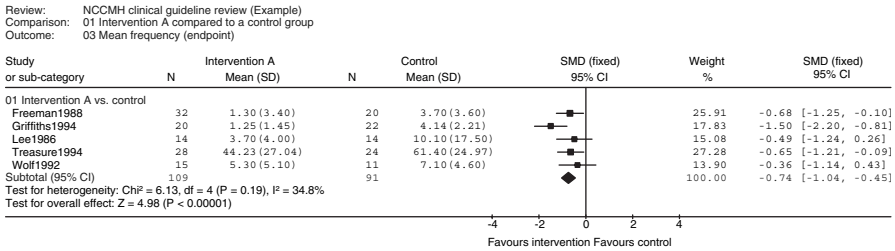
The CI shows with 95% certainty the range within which the true treatment effect should lie, and can be used to determine statistical significance. If the CI does not cross the ‘line of no effect’, the effect is statistically significant.

Continuous outcomes were analysed as weighted mean differences (WMD), or as standardised mean differences (SMD) when different measures were used in different studies to estimate the same underlying effect (for an example, see Figure 2). If provided, intention-to-treat data, using a method such as ‘last observation carried forward’ (LOCF), were preferred over data from completers.

**Figure 1: Example of a forest plot displaying dichotomous data**



**Figure 2: Example of a forest plot displaying continuous data**



To check for consistency between studies, both the  $I^2$  test of heterogeneity and a visual inspection of the forest plots were used. The  $I^2$  statistic describes the proportion of total variation in study estimates that is due to heterogeneity (Higgins & Thompson, 2002). The  $I^2$  statistic was interpreted in the following way:

- Greater than 50%: notable heterogeneity. (An attempt was made to explain the variation; for example, outliers were removed from the analysis or sub-analyses were conducted to examine the possibility of moderators. If studies with heterogeneous results were found to be comparable, a random-effects model was used to summarise the results [DerSimonian & Laird, 1986]. In the random effects analysis, heterogeneity is accounted for both in the width of CIs and in the estimate of the treatment effect. With decreasing heterogeneity, the random effects approach moves asymptotically towards a fixed-effects model.)
- 30 to 50%: moderate heterogeneity (both the chi-squared test of heterogeneity and a visual inspection of the forest plot were used to decide between a fixed- and random-effects model).
- Less than 30%: mild heterogeneity (a fixed-effects model was used to synthesise the results).

To explore the possibility that the results entered into each meta-analysis suffered from publication bias, data from included studies were entered, where there was sufficient data, into a funnel plot. Asymmetry of the plot was taken to indicate possible publication bias and investigated further.

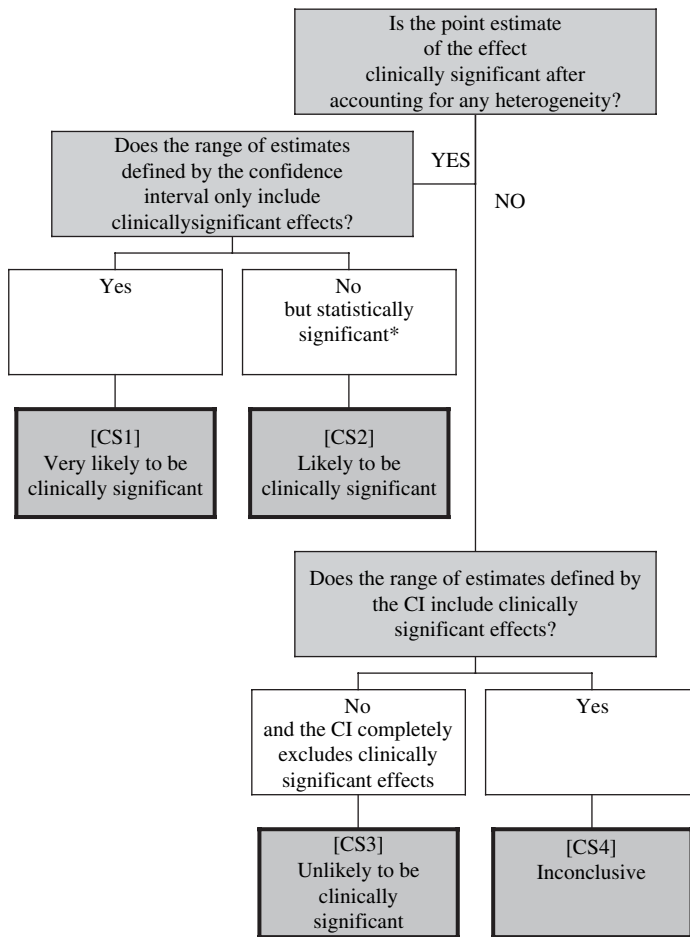
Forest plots included lines for studies that were believed to contain eligible data even if the data were missing from the analysis in the published study. An estimate of the proportion of eligible data that were missing (because some studies did not include all relevant outcomes) was calculated for each analysis.

### 3.5.5 Forming the clinical summaries and recommendations

Once the evidence profile tables relating to a particular clinical question were completed, summary tables incorporating important information from the evidence profile and an assessment of the clinical significance of the evidence were produced (these tables are presented in the evidence chapters). Finally, the systematic reviewer, in conjunction with the topic group lead, produced a clinical summary.

In order to facilitate consistency in generating and drafting the clinical summaries, a decision tree was used to help determine, for each comparison, the likelihood of the effect being clinically significant (see Figure 3). The decision tree was designed to be used as one step in the interpretation of the evidence (primarily to separate clinically

**Figure 3: Decision tree for helping to judge the likelihood of clinical significance**



\*Efficacy outcomes with large effect sizes and very wide CIs should be interpreted with caution and should be described as inconclusive (CS4), especially if there is only one small study.

### *Methods used to develop this guideline*

important from clinically negligible effects) and was not designed to replace clinical judgement. For each comparison, the GDG defined *a priori* a clinically significant threshold, taking into account both the comparison group and the outcome.

As shown in Figure 3, the review team first classified the point estimate of the effect as clinically significant or not. For example, if a relative risk of 0.75 was considered to be the threshold, then a point estimate of 0.73 (as can be seen in Figure 1), would meet the criteria for clinical significance. Where heterogeneity between studies was judged problematic, in the first instance an attempt was made to explain the cause of the heterogeneity (for example, outliers were removed from the analysis or sub-analyses were conducted to examine the possibility of moderators). Where homogeneity could not be achieved, a random-effects model was used.

Where the point estimate of the effect exceeded the threshold, a further consideration was made about the precision of the evidence by examining the range of estimates defined by the CI. Where the effect size was judged clinically significant for the full range of plausible estimates, the result was described as *very likely to be clinically significant* (that is CS1). In situations where the CI included clinically unimportant values, but the point estimate was both clinically and statistically significant, the result was described as *likely to be clinically significant* (that is CS2). However, if the CI crossed the line of no effect (that is, the result was not statistically significant), the result was described as *inconclusive* (that is CS4).

Where the point estimate did not meet the criteria for clinical significance and the CI completely excluded clinically significant values, the result was described as *unlikely to be clinically significant* (that is, CS3). Alternatively, if the CI included both clinically significant and clinically unimportant values, the result was described as *inconclusive* (that is, CS4). In all cases described as inconclusive, the GDG used clinical judgement to interpret the results.

Once the evidence profile tables and clinical summaries were finalised and agreed by the GDG, the associated recommendations were produced, taking into account the trade-off between the benefits and risks as well as other important factors. These included economic considerations, values of the development group and society, and the group's awareness of practical issues (Eccles *et al.*, 1998).

### **3.5.6 Method used to answer a clinical question in the absence of appropriately designed, high-quality research**

In the absence of RCTs (or high-quality research of a design appropriate to the clinical question), or where the GDG were of the opinion (on the basis of previous searches or their knowledge of the literature) that there were unlikely to be such evidence, an informal consensus process was adopted. This process focused on those questions that the GDG considered a priority.

#### *Informal consensus*

The starting point for the process of informal consensus was that a member of the topic group identified, with help from the systematic reviewer, a narrative review that

most directly addressed the clinical question. Where this was not possible, a brief review of the recent literature was initiated.

This existing narrative review or new review was used as a basis for beginning an iterative process to identify lower levels of evidence relevant to the clinical question and to lead to written statements for the guideline. The process involved a number of steps:

- A description of what is known about the issues concerning the clinical question was written by one of the topic group members.
- Evidence from the existing review or new review was then presented in narrative form to the GDG and further comments were sought about the evidence and its perceived relevance to the clinical question.
- Based on the feedback from the GDG, additional information was sought and added to the information collected. This may have included studies that did not directly address the clinical question but were thought to contain relevant data.
- If, during the course of preparing the report, a significant body of primary-level studies (of appropriate design to answer the question) were identified, a full systematic review was done.
- At this time, subject possibly to further reviews of the evidence, a series of statements that directly addressed the clinical question was developed.
- Following this, on occasions and as deemed appropriate by the development group, the report was then sent to appointed experts outside of the GDG for peer review and comment. The information from this process was then fed back to the GDG for further discussion of the statements.
- Recommendations were then developed.
- After this final stage of comment, the statements and recommendations were again reviewed and agreed upon by the GDG.

### **3.6 HEALTH ECONOMICS REVIEW STRATEGIES**

The aim of the health economics literature review was to contribute to the guideline development process by providing evidence on the economic burden of mental disorders in the antenatal and postnatal period as well as on the relative cost effectiveness of different preventive and treatment options covered in the guideline. Where available, relevant evidence was collected and assessed in order to help the decision-making process.

This process was based on a preliminary analysis of the clinical evidence and had two stages:

- identification of areas with likely major resource implications within the scope of the guideline
- systematic review of existing data on the economic burden of mental disorders in the antenatal and postnatal period and evidence on cost effectiveness of interventions aimed at prevention and management of such disorders.

In addition, in areas with likely major cost implications where relevant data did not already exist, primary economic analyses based on decision-analytic economic modelling were undertaken alongside the guideline development process, in order to provide cost-effectiveness evidence and assist decision making.

## *Methods used to develop this guideline*

### *3.6.1 Key economic issues*

The following economic issues relating to the epidemiology and the management of mental disorders in the antenatal and postnatal period were identified by the GDG in collaboration with the health economist as primary key issues that should be considered in the guideline:

- the global economic burden of mental disorders experienced by women during pregnancy and in their first postnatal year, with specific reference to the UK
- cost effectiveness of psychological interventions for the prevention and treatment of depression in the postnatal period
- cost effectiveness of specialist perinatal mental health services for the management of women with mental disorders in the antenatal and postnatal period.

### **3.6.2 Systematic literature review**

A systematic review of the health economics evidence was conducted. The aim of the review was threefold:

- to identify publications providing information on the economic burden of mental disorders during pregnancy and in the first postnatal year relevant to the UK context
- to identify existing economic evaluations of psychological interventions for the prevention and treatment of depression in the postnatal period, as well as of specialist perinatal mental health services for the management of women with mental disorders in the antenatal and postnatal period, that were transferable to the UK patient population and healthcare setting
- to identify studies reporting relevant health state utility data transferable to the UK population to facilitate a possible cost–utility modelling process.

Although no attempt was made to review systematically studies with only resource use or cost data, relevant UK-based information was extracted for future modelling exercises if it was considered appropriate.

### **3.6.3 Search strategy**

For the systematic review of economic evidence, the standard mental-health-related bibliographic databases (EMBASE, MEDLINE, CINAHL, PsychINFO and HTA) were searched. For these databases, a health economics search filter adapted from the Centre for Reviews and Dissemination (CRD) at the University of York was used in combination with a general filter for antenatal- and postnatal-related mental disorders. The subject filter employed a combination of free-text terms and medical subject headings, with subject headings having been exploded. Additional searches were performed in specific health economics databases (NHS EED, OHE HEED). HTA and NHS EED databases were accessed via the Cochrane Library, using the general filter for antenatal- and postnatal-related mental disorders. OHE HEED was searched using a shorter, database-specific strategy. Initial searches were performed between



February and March 2005. The searches were updated regularly, with the final search between 6 and 8 weeks before the first consultation. Search strategies used for the health economics systematic review are presented in Appendix 6.

In parallel to searches of electronic databases, reference lists of eligible studies and relevant reviews were searched by hand, and experts in the field of antenatal and postnatal mental health and mental health economics were contacted in order to identify additional relevant published and unpublished studies. Studies included in the clinical evidence review were also screened for economic evidence.

### **3.6.4 Review process**

The database searches for general health economics evidence for bipolar disorder resulted in 84 potentially eligible references. A further two possibly eligible references were found by hand searching. Full texts of all potentially eligible studies (including those for which relevance/eligibility was not clear from the abstract) were obtained. These publications were then assessed against a set of standard inclusion criteria by the health economist, and papers eligible for inclusion as economic evaluations were subsequently assessed for internal validity. The quality assessment was based on the 35-point checklist used by the *British Medical Journal* to assist referees in appraising full economic analyses (Drummond & Jefferson, 1996) (see Appendix 12).

### **3.6.5 Selection criteria**

The following inclusion criteria were applied to select studies identified by the economic searches for further analysis:

- No restriction was placed on language or publication status of the papers.
- Studies published between 1985 and 2006 were included. This date restriction was imposed in order to obtain data relevant to current healthcare settings and costs.
- Only studies from Organisation for Economic Cooperation and Development (OECD) countries were included, as the aim of the review was to identify economic information transferable to the UK context.
- Selection criteria based on types of clinical conditions and patients were identical to the clinical literature review (see Appendix 12).
- Studies were included provided that sufficient details regarding methods and results were available to enable the methodological quality of the study to be assessed and provided that the study's data and results were extractable.

Additional selection criteria were applied in the case of economic evaluations:

- Only full economic evaluations that compared two or more options and considered both costs and consequences (that is cost–minimisation analysis, cost–consequences analysis, cost–effectiveness analysis, cost–utility analysis or cost–benefit analysis) were included in the review.

### *Methods used to develop this guideline*

- Economic studies were considered only if they utilised clinical evidence derived from a meta-analysis, a well-conducted literature review, an RCT, a quasi-experimental trial or a cohort study.

#### **3.6.6 Data extraction**

Data were extracted by the health economist using an economic data extraction form (Appendix 13). Masked assessment, whereby data extractors are blind to the details of journal, authors, and so on, was not undertaken.

#### **3.6.7 Presentation of the results**

The economic evidence identified in the health economics systematic review is summarised in the respective chapters of the guideline, following presentation of the clinical evidence. Results of additional economic modelling undertaken alongside the guideline development process are also presented in the relevant chapters.

### **3.7 STAKEHOLDER CONTRIBUTIONS**

Professionals, service users and companies have contributed to and commented on the guideline at key stages in its development. Stakeholders for this guideline include:

- service user/carer stakeholders: the national service user and carer organisations that represent people whose care is described in this guideline
- professional stakeholders: the national organisations that represent healthcare professionals who are providing services to service users
- commercial stakeholders: the companies that manufacture medicines used in the treatment of mental disorders
- PCTs
- DH and Welsh Assembly Government.

Stakeholders have been involved in the guideline's development at the following points:

- commenting on the initial scope of the guideline and attending a briefing meeting held by NICE
- contributing possible clinical questions and lists of evidence to the GDG
- commenting on the first and second drafts of the guideline.

### **3.8 TESTIMONIES FROM WOMEN WITH MENTAL DISORDERS IN THE ANTENATAL AND POSTNATAL PERIOD**

Throughout this document, there are illustrations of women's experiences of mental health problems, treatment and services in the antenatal and postnatal periods; these

are in the form of short vignettes or longer testimonies. The intention behind the use of these extracts is to add to the understanding of individual experience described in this guideline.

The writers of the testimonies and vignettes were contacted primarily through service user and carer stakeholder organisations. They were asked to consider the following questions:

- If you had experienced mental health problems at some time before you became pregnant, did you discuss this at any point with healthcare professionals (GP, midwife and so on)? What information were you given about either starting or continuing treatment for this problem through your pregnancy and after birth?
- If your first experience of mental health problems occurred during pregnancy or within a year after giving birth, when and how did you first become aware that you had a mental health problem?
- What possible treatments were discussed with you and what treatments did you receive? Did the treatment help you feel better? (Please describe what worked for you and what didn't work for you).
- Did you attend a support group and was this helpful?
- How would you describe your relationship with your healthcare professional(s) (GP/midwife/health visitor/CPN/psychiatrist, obstetrician and so on)?
- How do you feel now?
- In what ways has your experience of mental health problems during the antenatal and postnatal period affected your life and the lives of those close to you?

Each writer of a testimony or vignette was also asked to sign a consent form to allow use of the material in the guideline.

### **3.9 VALIDATION OF THIS GUIDELINE**

Registered stakeholders had two opportunities to comment on the draft guideline, which was posted on the NICE website during the consultation period. The GRP also reviewed the guideline and checked that stakeholders' comments had been addressed.

Following the consultation period, the GDG finalised the recommendations and the NCCMH produced the final documents. These were then submitted to NICE. NICE then formally approved the guideline and issued its guidance to the NHS in England and Wales.